Foundation of Cryptography (0368-4162-01), Lecture 3

Hardcore Predicates for Any One-way Function

Iftach Haitner, Tel Aviv University

November 22, 2011

Hardcore Predicates

Definition 1 (hardcore predicates)

A polynomial-time computable function $b : \{0,1\}^n \mapsto \{0,1\}$ is an hardcore predicate of the function $f \colon \{0,1\}^n \mapsto \{0,1\}^n$, if

$$\Pr_{x \leftarrow \{0,1\}^n}[\mathsf{P}(f(x)) = b(x)] \leq \frac{1}{2} + \operatorname{neg}(n),$$

for any PPT P.

Theorem 2 (Goldreich-Levin)

Let $f \colon \{0,1\}^n \mapsto \{0,1\}^n$ be a OWF, and define $g \colon \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}^n \times \{0,1\}^n$ as $g(x,r) = f(x), r$. Then $b(x,r) = \langle x, r \rangle_2$, is an hardcore predicate of g.

Note that if f is one-to-one, then so is g.

Section 1

The Information Theoretic Case

Definition 3 (min-entropy)

The min entropy of a random variable X, is defined

$$\mathrm{H}_{\infty}(X) := \min_{y \in \mathrm{Supp}(X)} \log \frac{1}{\Pr_X[y]}.$$

Examples

- X is uniform over a set of size 2^k
- $(X \mid f(X) = y)$, where $f \colon \{0,1\}^n \mapsto \{0,1\}^n$ is 2^k to 1 and X is uniform over $\{0,1\}^n$

Pairwise independent hashing

Definition 4 (pairwise independent hash functions)

A function family $\mathcal{H}$ from $\{0,1\}^n$ to $\{0,1\}^m$ is pairwise independent, if for every $x \neq x' \in \{0,1\}^n$ and $y, y' \in \{0,1\}^m$, it holds that $\Pr_{h \leftarrow \mathcal{H}}[h(x) = y \wedge h(x') = y')] = 2^{-2m}$.

Lemma 5 (leftover hash lemma)

Let X be a random variable over $\{0,1\}^n$ with $\mathrm{H}_\infty(X) \geq k$ and let $\mathcal{H}$ be a family of pairwise independent hash functions from $\{0,1\}^n$ to $\{0,1\}^m$, then

$$\mathrm{SD}((h, h(x))_{h \leftarrow \mathcal{H}, x \leftarrow X}, (h, y)_{h \leftarrow \mathcal{H}, y \leftarrow \{0,1\}^m}) \leq 2^{(m-k-2))/2}.$$

* We typically simply write $\mathrm{SD}((H, H(X)), (H, U_m))$, where H is uniformly distributed over $\mathcal{H}$.

Efficient function families

Definition 6 (efficient function family)

An ensemble of function families $\mathcal{F} = \{\mathcal{F}_n\}_{n \in \mathbb{N}}$ is efficient, if the following hold:

Samplable. $\mathcal{F}$ is samplable in polynomial-time: there exists a PPT that given 1^n, outputs (the description of) a uniform element in $\mathcal{F}_n$.

Efficient. There exists a polynomial-time algorithm that given $x \in \{0,1\}^n$ and (a description of) $f \in \mathcal{F}_n$, outputs $f(x)$.

Hardcore predicate for regular OWF

Lemma 7

Let $f:\{0,1\}^n \mapsto \{0,1\}^n$ be a $d(n) \in 2^{\omega(\log n)}$ regular function and let $\mathcal{H} = \{\mathcal{H}_n\}$ be an efficient family of Boolean pairwise independent hash functions over $\{0,1\}^n$. Define $g\colon \{0,1\}^n \times \mathcal{H}_n \mapsto \{0,1\}^n \times \mathcal{H}_n$ as

$$g(x,h) = (f(x), h),$$

then $b(x,h) = h(x)$ is an hardcore predicate of g.

How does it relate to the computational case?

Proof: We prove the claim by showing that

Claim 8

$\mathrm{SD}\left((f(U_n), H, H(U_n)), (f(U_n), H, U_1)\right) = \mathrm{neg}(n)$, where the rv $H = H(n)$ is uniformly distributed over $\mathcal{H}_n$.

Does this conclude the proof?

Proving Claim 8

Proof: For $y \in f(\{0,1\}^n) := \{f(x) \colon x \in \{0,1\}^n\}$, let the rv X_y be uniformly distributed over $f^{-1}(y) := \{x \in \{0,1\}^n \colon f(x) = y\}$.

$$
\begin{aligned}
&\mathrm{SD}((f(U_n), H, H(U_n)), (f(U_n), H, U_1)) \\
&= \sum_{y \in f(\{0,1\}^n)} \Pr[f(U_n) = y] \cdot \mathrm{SD}\big((f(U_n), H, H(U_n) \mid f(U_n) = y) \\
&\qquad\qquad\qquad , (f(U_n), H, U_1 \mid f(U_n) = y)\big) \\
&= \sum_{y \in f(\{0,1\}^n)} \Pr[f(U_n) = y] \cdot \mathrm{SD}\left((y, H, H(X_y)), (y, H, U_1)\right) \\
&\leq \max_{y \in f(\{0,1\}^n)} \mathrm{SD}((y, H, H(X_y)), (y, H, U_1)) \\
&\leq \max_{y \in f(\{0,1\}^n)} \mathrm{SD}((H, H(X_y)), (H, U_1))
\end{aligned}
$$

Since $\mathrm{H}_\infty(X_y) = \log(d(n))$ for any $y \in f(\{0,1\}^n)$,
The leftover hash lemma yields that

$$\begin{aligned} \mathrm{SD}((H, H(X_y)), (H, U_1)) &\leq 2^{(1-\mathrm{H}_\infty(X_y)-2))/2} \\ &= 2^{(1-\log(d(n)))/2} = \mathrm{neg}(n). \qquad \square \end{aligned}$$

Further remarks

Remark 9

- We can output $\Theta(\log d(n))$ bits,
- g and b are not defined over all input length.

Section 2

The Computational Case

Proving Goldreich-Levin Theorem

Theorem 10 (Goldreich-Levin)

Let $f\colon \{0,1\}^n \mapsto \{0,1\}^n$ be a OWF, and define $g\colon \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}^n \times \{0,1\}^n$ as $g(x,r) = f(x), r$. Then $b(x,r) = \langle x, r\rangle_2$, is an hardcore predicate of g.

Note that if $b(x,r)$ is (almost) a family of pairwise independent hash functions.

Proof: Assume $\exists$ PPT A, $p \in$ poly and infinite set $\mathcal{I} \subseteq \mathbb{N}$ with

$$\Pr[\mathrm{A}(g(U_n, R_n)) = b(U_n, R_n)] \geq \frac{1}{2} + \frac{1}{p(n)}, \tag{1}$$

for any $n \in \mathcal{I}$, where U_n and R_n are uniformly (and independently) distributed over $\{0,1\}^n$.

We show $\exists$ PPT B and $p' \in$ poly with

$$\Pr_{y \leftarrow f(U_n)}[\mathrm{B}(y) \in f^{-1}(y) \geq \frac{1}{p'(n)}, \tag{2}$$

for every $n \in \mathcal{I}$. In the following fix $n \in \mathcal{I}$.

Focusing on a good set

Claim 11

There exists a set $\mathcal{S} \subseteq \{0,1\}^n$ with

1. $\frac{|\mathcal{S}|}{2^n} \geq \frac{1}{2p(n)}$, and
2. $\alpha(x) := \Pr[\mathsf{A}(f(x), R_n) = b(x, R_n)] \geq \frac{1}{2} + \frac{1}{2p(n)}$, $\forall x \in S$.

Proof: Let $\mathcal{S} := \{x \in \{0,1\}^n \colon \alpha(x) \geq \frac{1}{2} + \frac{1}{2p(n)}\}$. It follows that

$$\begin{aligned}\Pr[\mathsf{A}(g(U_n, R_n)) = b(U_n, R_n)] &\leq \Pr[U_n \notin \mathcal{S}] \cdot \left(\frac{1}{2} + \frac{1}{2p(n)}\right) + \Pr[U_n \in \mathcal{S}] \\ &\leq \left(\frac{1}{2} + \frac{1}{2p(n)}\right) + \Pr[U_n \in \mathcal{S}] \square\end{aligned}$$

We will present $q \in$ poly and a PPT B such that

$$\Pr[\mathsf{B}(y = f(x)) \in f^{-1}(y) \geq \frac{1}{q(n)}, \tag{3}$$

for every $x \in \mathcal{S}$. Fix $x \in \mathcal{S}$.

The perfect case $\alpha(x) = 1$

For every $i \in [n]$, it holds that

$$\mathsf{A}(f(x), e^i) = b(x, e^i),$$

where $e^i = (\underbrace{0, \ldots, 0}_{i-1}, 1, \underbrace{0, \ldots, 0}_{n-i})$.

- Hence, $x_i = \langle x, e^i \rangle_2 = \mathsf{A}(f(x), e^i)$

We let $\mathsf{B}(f(x)) = (\mathsf{A}(f(x), e^1), \ldots, \mathsf{A}(f(x), e^n))$

Easy case: $\alpha(x) \geq 1 - \mathrm{neg}(n)$

Fact 12

1. *$\forall r \in \{0,1\}^n$, the rv $(r \oplus R_n)$ is uniformly dist. over $\{0,1\}^n$*
2. *$\forall w, y \in \{0,1\}^n$, it holds that $b(x,w) \oplus b(x,y) = b(x, w \oplus y)$*

Hence, $\forall i \in [n]$:

1. $\forall r \in \{0,1\}^n$ it holds that $x_i = b(x,r) \oplus b(x, r \oplus e^i)$
2. $\Pr[\mathrm{A}(f(x), R_n) = b(x, R_n) \wedge \mathrm{A}(f(x), R_n \oplus e^i) = b(x, R_n \oplus e^i)] \geq 1 - \mathrm{neg}(n)$

We let $\mathrm{B}(f(x)) = (\mathrm{A}(f(x), R_n) \oplus \mathrm{A}(f(x), R_n \oplus e^1)), \ldots, \mathrm{A}(f(x), R_n) \oplus \mathrm{A}(f(x), R_n \oplus e^n))$.

Intermediate case: $\alpha(x) \geq \frac{3}{4} + \frac{1}{q(n)}$

For any $i \in [n]$, it holds that

$$\begin{aligned}
&\Pr[A(f(x), R_n) \oplus A(f(x), R_n \oplus e^i) = x_i] \qquad (4)\\
&\geq \Pr[A(f(x), R_n) = b(x, R_n) \wedge A(f(x), R_n \oplus e^i) = b(x, R_n \oplus e^i)]\\
&\geq \frac{1}{2} + \frac{2}{q(n)}
\end{aligned}$$

Algorithm 13 (B)

Input: $f(x) \in \{0,1\}^n$

1. For every $i \in [n]$
   - Sample $r^1, \ldots, r^v \in \{0,1\}^n$ uniformly at random
   - Let $m_i = \mathrm{maj}_{j \in [v]}\{(A(f(x), r^j) \oplus A(f(x), r^j \oplus e^i)\}$
2. Output $(m_1, \ldots, m_n)$

B's success provability

The following holds for "large enough" $v = v(n)$.

Claim 14

For every $i \in [n]$, it holds that $\Pr[m_i = x_i] \geq 1 - \mathrm{neg}(n)$.

Proof: For $j \in [v]$, let the indicator rv W^j be 1, iif $A(f(x), r^j) \oplus A(f(x), r^j \oplus e^i) = x_i$.
We want to lowerbound $\Pr\left[\sum_{j=1}^{v} W^j > \frac{v}{2}\right]$.

- The W^j are iids and $\mathrm{E}[W^j] \geq \frac{1}{2} + \frac{2}{q(n)}$, for every $j \in [v]$

Lemma 15 (Hoeffding's inequality)

Let $X^1, \ldots, X^v$ be iid over $[0,1]$ with expectation μ. Then, $\Pr\left[\left|\frac{\sum_{j=i}^{v} X^j}{v} - \mu\right| \geq \varepsilon\right] \leq 2 \cdot \exp(-2\varepsilon^2 v)$ *for every* $\varepsilon > 0$.

We complete the proof taking $X^j = W^j$, $\varepsilon = 1/4q(n)$ and $v \in \omega(\log(n) \cdot q(n)^2)$.

The actual case: $\alpha(x) \geq \frac{1}{2} + \frac{1}{q(n)}$

- What goes wrong?
- Idea: guess the values of $\{b(x, r^1), \ldots, b(x, r^v)\}$ (instead of calling $\{\mathrm{A}(f(x), r^1), \ldots, \mathrm{A}(f(x), r^v)\}$)
- Problem: negligible success probability
- Solution: choose the samples in a *correlated* manner

Algorithm B

Fix $\ell = \ell(n)$ (will be $O(\log n)$) and set $v = 2^\ell - 1$.
We let $\mathcal{L} \subseteq [\ell]$ stands for non-empty subset.

Algorithm 16 (B)

Input: $f(x) \in \{0,1\}^n$

1. Sample uniformly (and independently) $t_1, \ldots, t_\ell \in \{0,1\}^n$
2. For all $\mathcal{L} \subseteq [\ell]$, set $r^{\mathcal{L}} = \bigoplus_{i \in \mathcal{L}} t^i$
3. Guess $\{b(x, t^i)\}$, and compute $\{b(x, r^{\mathcal{L}})\}$ (how?)
4. For all $i \in [n]$, let
   $m_i = \mathrm{maj}_{\mathcal{L} \subseteq \{0,1\}^n} \{\mathsf{A}(f(x), r^{\mathcal{L}} \oplus e^i) \oplus b(x, r^{\mathcal{L}})\}$
5. Output $(m_1, \ldots, m_n)$

Fix $i \in [n]$, and let $W^{\mathcal{L}}$ be 1, iff $\mathsf{A}(f(x), r^{\mathcal{L}} \oplus e^i) \oplus b(x, r^{\mathcal{L}}) = x_i$.
We want to lowerbound $\Pr[\sum_{\mathcal{L} \subseteq [\ell]} W^{\mathcal{L}} > \frac{v}{2}]$
Problem: the $W^{\mathcal{L}}$'s are *dependent*!

Analyzing B's success probability

1. Let $T^1, \ldots, T^\ell$ be iid over $\{0,1\}^n$.
2. For every $\mathcal{L} \subseteq [\ell]$, let $R^{\mathcal{L}} = \bigoplus_{i \in \mathcal{L}} T^i$.

Fact 17

1. *$\forall \mathcal{L} \subseteq [\ell]$, $R^{\mathcal{L}}$ is uniformly distributed over $\{0,1\}^n$*
2. *$\forall w, y \in \{0,1\}^n$ and $\forall \mathcal{L} \neq \mathcal{L}' \subseteq [\ell]$, it holds that* $\Pr[R^{\mathcal{L}} = w \wedge R^{\mathcal{L}'} = y] = \Pr[R^{\mathcal{L}} = w] \cdot \Pr[R^{\mathcal{L}'} = y]$

That is, the $R^{\mathcal{L}}$'s are *pairwise independent*.

Proving Fact 17(2)

Assume wlg. that $1 \in (\mathcal{L}' \setminus \mathcal{L})$.

$$
\begin{aligned}
&\Pr[R^{\mathcal{L}} = w \wedge R^{\mathcal{L}'} = y] \\
&= \sum_{(t^2,\ldots,t^\ell)\in\{0,1\}^{(\ell-1)n}} \Pr[(T^2,\ldots,T^\ell) = (t^2,\ldots,t^\ell)] \cdot \\
&\qquad \Pr[R^{\mathcal{L}} = w \wedge R^{\mathcal{L}'} = y \mid (T^2,\ldots,T^\ell) = (t^2,\ldots,t^\ell)] \\
&= \sum_{(t^2,\ldots,t^\ell):\, (\bigoplus_{i\in\mathcal{L}} t^i) = w} \Pr[(T^2,\ldots,T^\ell) = (t^2,\ldots,t^\ell)] \\
&\qquad \cdot \Pr[R^{\mathcal{L}} = w \wedge R^{\mathcal{L}'} = y \mid (T^2,\ldots,T^\ell) = (t^2,\ldots,t^\ell)] \\
&= \sum_{(t^2,\ldots,t^\ell):\, (\bigoplus_{i\in\mathcal{L}} t^i) = w} \Pr[(T^2,\ldots,T^\ell) = (t^2,\ldots,t^\ell)] \cdot 2^{-n} \\
&= 2^{-n} \cdot 2^{-n} = \Pr[R^{\mathcal{L}} = w] \cdot \Pr[R^{\mathcal{L}'} = y] \square
\end{aligned}
$$

Pairwise independence variables

Definition 18 (pairwise independent random variables)

A sequence of random variables $X^1, \ldots, X^v$ is pairwise independent, if $\forall i \neq j \in [v]$ and $\forall a, b$, it holds that

$$\Pr[X^i = a \wedge X^j = b] = \Pr[X^i = a] \cdot \Pr[X^j = b]$$

For every $\mathcal{L} \neq \mathcal{L}' \subseteq [\ell]$, the rvs $R^{\mathcal{L}}$ and $R^{\mathcal{L}'}$ are pairwise independent, and therefore also $W^{\mathcal{L}}$ and $W^{\mathcal{L}'}$ (why?).

Lemma 19 (Chebyshev's inequality)

Let $X^1, \ldots, X^v$ be pairwise-independent random variables with expectation μ and variance σ^2. Then, for every $\varepsilon > 0$,

$$\Pr\left[\left|\frac{\sum_{j=1}^{v} X^j}{v} - \mu\right| \geq \varepsilon\right] \leq \frac{\sigma^2}{\varepsilon^2 v}$$

B's success provability cont

Assuming that B always guesses $\{b(x, t^i)\}$ correctly, then for every $\mathcal{L} \subseteq [\ell]$

- $\mathrm{E}[W^{\mathcal{L}}] \geq \frac{1}{2} + \frac{1}{q(n)}$
- $\mathrm{Var}(W^{\mathcal{L}}) := \mathrm{E}[W^{\mathcal{L}}]^2 - \mathrm{E}[(W^{\mathcal{L}})^2] \leq 1$

Taking $\varepsilon = 1/2q(n)$ and $v = 2n/\varepsilon^2$ (i.e., $\ell = \lceil \log(2n/\varepsilon^2) \rceil$), yields that

$$\Pr[m_i = x_i] = \Pr\left[\frac{\sum_{\mathcal{L} \subseteq [\ell]} W^{\mathcal{L}}}{v} > \frac{1}{2}\right] \geq 1 - \frac{1}{2n} \tag{5}$$

and by a union bound, B outputs x with probability $\frac{1}{2}$.
Taking the guessing into account, yields that B outputs x with probability at least $2^{-\ell-1} \in \Omega(n/q(n)^2)$.

Reflections

Hardcore functions. Similar ideas allows to output $\log n$ "pseudorandom bits"

Alternative proof for the LHL. Let X be a rv with over $\{0,1\}^n$ with $\mathrm{H}_\infty(X) \geq t$, and assume that $\mathrm{SD}((R_n, \langle R_n, X\rangle_2), (R_n, U_1)) > \alpha = 2^{-c \cdot t}$ for some universal $c > 0$. Hence

1. $\exists$ (a possibly inefficient) algorithm D that distinguishes $(R_n, \langle R_n, X\rangle_2)$ from (R_n, U_1) with advantage α
2. $\exists$A that predicts $\langle R_n, X\rangle_2$ given R_n with prob $\frac{1}{2} + \alpha$
3. (by GL) $\exists$B that guesses X "from nothing", with prob $\alpha^{O(1)} > 2^{-t}$

Reflections cont.

List decoding. An efficient encoding $C\colon \{0,1\}^n \mapsto \{0,1\}^m$, and a decoder D. Such that the following holds for any $x \in \{0,1\}^n$ and c of hamming distance $\frac{1}{2} - \delta$ from $C(x)$:
$D(c, \delta)$ outputs a list of size at most $\text{poly}(1/\delta)$ that whp. contains x
The code we used here is known as the *Hadamard* code

LPN - learning parity with noise. Find x given polynomially many samples of $\langle x, R_n \rangle_2 + N$, where $\Pr[N = 1] \leq \frac{1}{2} - \delta$.
The difference comparing to Goldreich-Levin – no control over the R_n's.